REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY) 03-12-2015	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 10-Mar-2014 - 9-Sep-2015

4. TITLE AND SUBTITLE Final Report: Acquisition of a High Performance Computing Instrument for Big Data Research and Education	5a. CONTRACT NUMBER W911NF-14-1-0119
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 206022
6. AUTHORS Justin Zhan, Huiming Yu, Clay Gloster, Rajeev Agrawal, Christopher C. Doss	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES North Carolina A&T State University 1601 East Market Street Greensboro, NC 27411 -0001	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 64728-NS-REP.3

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
   The objective of this proposal is to fund the purchase of an integrated high performance computing (HPC) instrument for big data research and education at North Carolina Agricultural and Technical State University (NCA&T). This instrument will promote innovation and translational big data research at NCA&T. It will make NCA&T (a land grant HBCU) a strong player in big data research and education. The research groups involved in this proposal consist of PI Zhan from the Department of Computer Science and other key personnel from the Departments of Computer Science, Computer Systems Technology, and Electrical and Computer Engineering at

15. SUBJECT TERMS
Final report, high performance computing, Big Data research

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Kaushik Roy
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	UU		19b. TELEPHONE NUMBER 336-285-3691

Final Report: Acquisition of a High Performance Computing Instrument for Big Data Research and Education

ABSTRACT

The objective of this proposal is to fund the purchase of an integrated high performance computing (HPC) instrument for big data research and education at North Carolina Agricultural and Technical State University (NCA&T). This instrument will promote innovation and translational big data research at NCA&T. It will make NCA&T (a land grant HBCU) a strong player in big data research and education. The research groups involved in this proposal consist of PI Zhan from the Department of Computer Science and other key personnel from the Departments of Computer Science, Computer Systems Technology, and Electrical and Computer Engineering at NCA&T. They have been actively conducting research on big data analytics that require such an instrument for mass data processing and storage. The HPC instrument will be a critical asset for the research groups involving big data analysis and management. It will advance various big data research and educational projects at NCA&T. The instrument will have a signi?cant and lasting broader impact on basic data science, provide new opportunities at NCA&T, and facilitate the advancement of outstanding next-generation faculty members working across the disciplines.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

<u>Received</u>   <u>Paper</u>

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

<u>Received</u>   <u>Paper</u>

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

• Albu-Shamah, A, Zhan, J. Obesity in Online Social Networks, The Sixth ASE International Conference on Social Computing, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Meyer, B., Zhan, J., A Look Inside the Nodes Cohesive Attribute Micro-Clustering, The Second ASE International Conference on Big Data Science and Computing, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Selim, H., Chopade, P., Zhan, J., Statistical Modeling and Scalable, Interactive Visualization of Large Scale Big Data Networks, The Second ASE International Conference on Big Data Science and Computing, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Selim, H., Chopade, P., Zhan, J., Structural Analysis and Interactive Visualization of Large Scale Big Data Networks, The IEEE 11th International Conference and Expo on Emerging Technologies for a Smarter World, Stony Brook University, New York, October 29-30, 2014.

• Kaur, B., Zhan, J. Malware Detection using the Weak Estimator, The Third ASE International Conference on Cyber Security, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Kaur, B., Blow, M., Zhan, J., Authenticity of Images in Social Media, NSF Workshop on Big Data Security and Privacy, University of Texas at Dallas, TX, September 16-17, 2014.

• Chopade, P., Zhan, J., Community Detection in Large Scale Big Data Networks, The Second ASE International Conference on Big Data, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Selim, H., Chopade, P., Zhan, J., Statistical Modeling and Scalable, Interactive Visualization of Large Scale/Big Data Networks, ASE International Workshop on Big Data Analytics for Predictive Organization and Big Transformations, Stanford University, CA, USA, May 27, 2014.

• Chopade, P., Bikdash, M., Zhan, J., Grid reliability: Need of developing self-healing and self-mitigating Smart Power Grid, The Second International Symposium on Energy Challenges and Mechanics (ECM2),19-21 August 2014, Aberdeen, Scotland, UK.

• X. Fang and J. Zhan, A Novel Approach for Evaluating Semantic Similarity Measures, The Sixth ASE International Conference on Social Computing, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Lu, Y., Fang, X., Zhan, J., Towards Data Readiness Level For Structured Data, The Second ASE International Conference on Big Data, Stanford University, Stanford, CA, USA, May 27-31, 2014.

• Lu, Y., Fang, X., Zhan, J., Data Readiness Level For Unstructured Data with a Focus on Unindexed Data, The Third ASE International Conference on Big Data, Tsinghua University, Beijing, China, August 4-7, 2014 (Best Paper Award).

• Yi-Ting Chiang, Tsan-Sheng Hsu, Churn-Jung Liau, Yun-Ching Liu, Chih-Hao Shen, Da-Wei Wang and Justin Zhan, An Information-Theoretic Approach for Secure Protocol Composition, The Tenth International Conference on Security and Privacy in Communication Networks, Beijing, China, September 24-26, 2014.

• Pravin Chopade, Kenneth Flurchick, Justin Zhan and Marwan Bikdash, Visualization Techniques for Large-Scale Big Data Networks: Smart Power Grid Survivability in a Complex Operating Environment, The Fourth ASE International Conference on Big Data, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

• Pravin Chopade, Justin Zhan, and Marwan Bikdash, Supervized Community Detection for Big Data and Large-Scale Complex Networks, The Fourth ASE International Conference on Big Data, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

• Bradley Meyer and Justin Zhan, Drowning in Opinions: Extracting the Pearls, The Third ASE International Conference on Social Informatics, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

• Haysam Selim, Pravin Chopade, and Justin Zhan, Node Degree and Edge Clustering Correlation for Community Detection in Big Data and Large-Scale Networks, The Fourth ASE International Conference on Big Data, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

• Ahmad Albu-Shamah and Justin Zhan, Discovering Hidden Networks Based On Twitter Texts, The Third ASE International Conference on Social Informatics, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

• Balkirat Kaur, Malcolm Blow, and Justin Zhan, Digital Image Authentication in Social Media, The Sixth ASE International Conference on Privacy, Security, Risk, and Trust, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

• Yogeshwar Rao Bachupally and Justin Zhan, Towards a Recommender System for DataBridge, The Fourth ASE International Conference on Big Data, Harvard University, Cambridge, MA, USA, December 14-16, 2014.

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

<u>Received</u>　　　　<u>Paper</u>

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

<u>Received</u>　　　　<u>Paper</u>

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

<u>Received</u>　　　　<u>Paper</u>

12/03/2015　2.00　Pravin, Chopade, Justin, Zhan . Structural and Functional Analytics for Community Detection in Large-Scale Complex Networks,
Journal of Big Data (09 2014)

12/03/2015　1.00　Xing, Fang, Justin , Zhan . Sentiment Analysis Using Product Review Data,
Journal of Big Data (09 2014)

TOTAL:　　**2**

Number of Manuscripts:

Books

<u>Received</u>   <u>Book</u>

TOTAL:

<u>Received</u>   <u>Book Chapter</u>

TOTAL:

Patents Submitted

Patents Awarded

Awards

- Outstanding Research Award at the North Carolina A&T State University, April 2014.
- Best Paper Award at the 2014 Third International Conference on Big Data Science and Computing, August 2014.

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Bradley Meyer	1.00	
Xing Fang	1.00	
FTE Equivalent:	**2.00**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Pravin Chopade	1.00
FTE Equivalent:	**1.00**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Justin Zhan	0.00	
Clay Gloster	0.00	
Huiming Yu	0.00	
Rajeev Agrawal	0.00	
Christopher C. Doss	0.00	
FTE Equivalent:	**0.00**	
Total Number:	**5**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics
This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Impact on Research Infrastructure

Adding the ability to perform Big Data research and education will provide an invaluable asset to various basic research activities on-going at NCA&T. The research groups involved in this proposal consist of researchers from the Departments of Computer Science, Computer Systems Technology, and Electrical and Computer Engineering at NCA&T. They have been actively conducting research on high performance computing and reconfigurable computing. All of these research activities require a high performance computing instrument for mass data processing and storage. The high performance computing instrument will be a critical asset for the research groups involving Big Data analysis and management. It will advance various Big Data research and educational projects at NCA&T.

The integrated research instrument described in this proposal, if funded and purchased, will have a significant and lasting broader impact on basic data science; provide new opportunities at NCA&T; and facilitate the advancement of outstanding next-generation faculty members working across the disciplines. As a leading producer of African-American baccalaureate, masters, and doctoral degree-holding engineers, this integrated research instrument will position NCA&T to be a strong player in Big Data research and education.

Impact on Education and Student Participation

This instrument will strengthen educational training activities by enabling innovative cross-disciplinary courses and promoting our education outreach activity with K-12 teachers and students. A key use of the HPC will be to provide training in using the advanced computing environments (hardware and software) and support for faculty and students across the university. NCA&T has considerable expertise in computing sciences including computer science, computer system technologies, and electrical and computer engineering. These are well-established engineering programs with an increasing focus in the areas of computing sciences. The requested instrument will be used to continue to provide this training and education and will benefit the following courses:

• COMP790 - Big Data; COMP690 - Social Computing; COMP790 - Network Science; COMP590- Parallel/Hybrid Programming; COMP590 - Data Fusion
• CSE629 - Introduction to Computational Science; CSE702 - Comprehensive Numerical Analysis; CSE703 - Software Principles and Programming in Scalable Parallel Computing; CSE704 - Computational Modeling and Visualization; CSE713 - Multi-Scale Physics and Modeling; CSE711 - Nano-science and Nano-engineering; CSE712 - Nano-Scale Technology; CSE713 - Multi-Scale & Multi-Physics Modeling; CSE803 - Advanced High Performance and Scalable Computing; CSE804 - Advanced Scientific Visualization; CSE805 - Visual Analytics and Data Mining
• ECT 414 - ASIC/FPGA Design; ECT 432 Computer Systems Architecture; ITT 433 - Intro to High Performance Computing; ITT 434 - HPC architecture and system administration; ITT 435 - Intro to Parallel Programming; ITT 615 - Network Security Applications; ITT 725: Wide Area Networks; ITT 731 - Introduction to Knowledge Discovery in Databases; ITT 747 - Secure Wireless and Wired Data Networks; ITT 750 - Computer System Security; ITT 752 - Advanced Computer Forensics.
• ECEN 423 Digital Systems I; ECEN 429 Digital Systems Lab; ECEN 623 Advanced Digital Systems; ECEN 624 Computer Architecture; ECEN 723 System Design with Programmable Logic Devices; ECEN 821 Advanced Computer Architecture; ECEN 867 Neural Networks Design.

Educational activities also include research training for students in the areas of GPU and FPGA computational implementations for these computationally intensive Big Data application areas, and large-scale, distributed modeling developments and applications.

Impact on Local High School Teachers/Students

We plan to offer a week-long on-campus Bridge to Big Data summer camp each year for high school students and teachers to participate in DoD funded projects. The camp will mainly target junior and senior high school students and teachers. We plan to recruit about 50 students and teachers for participation in the camp each year. The investigators will give seminars and demonstrations about bridging Big Data from various disciplinary perspectives during this camp to attract students to the field of computing sciences. The demonstrations will not only excite high school students but also educate them about the Big Data problems facing the digital age.

Impact on Army and National Defense

The work will benefit army and national defense. Since 9/11, the amount of data from drones and other surveillance technology has risen 1,600 percent. Our armed services now have approximately 7 million computing devices–a number that is expected to double by 2020. There is no shortage of data, but there is a dearth of analysis. The challenges and opportunities created by big data are by now well documented for a range of fields, but managing the growing amount of available information has never been more relevant to how our country is fighting wars and planning for future threats. On the battlefield, the time it takes to access intelligence can be a matter of life or death. Harvesting, analyzing, and rapidly converting large data sets into actionable intelligence are currency for the military. This research will add significant values for army and national defense.

The widespread use of unmanned autonomous surveillance systems in military operations offers significant advanced new capability while simultaneously introducing complex challenges in the rapid development/deployment and testing of these systems. Field commanders incorporating these systems into their tactical maneuvers typically impose additional requirements and specifications. The solution to many of these issues is to use Field Programmable Gate Arrays (FPGAs) in conjunction with onboard processors as a platform to develop a context neutral payload that supports a myriad of sensors types, provides

superior processing power, and can reduce the data requirements for communication of sensor data to satellites or ground stations. Reconfigurable computing systems or systems that combine onboard processors with one or more FPGAs offer programmable hardware systems that can be reloaded for disparate applications while using the same hardware.

Technology Transfer